Forum:	The Group of Twenty (G20)
Issue:	Implementation of Big Data infrastructure in the financial sector
Student Officer:	Nickolas Pachis
Position:	Deputy President

PERSONAL INTRODUCTION

Dear delegates of G20,

My name is Nickolas Pachis and I am in IB1 at Platon School. I am extremely honored to be serving as one of the two co-chairs of the G20 in this year's Platon School MUN Conference.

This topic is relatively new in the grand scheme of things. Data, and by extension, Big Data, plays a key role in the development of the modern world. It is a multi-faceted and complex topic, which means that this guide only scratches the surface of almost every aspect of the issue at hand. It is expected that you do your own research using the sources in the Bibliography section or by collecting information from sources you deem credible. Another important point that is integral to the correct implementation of Big Data is privacy and the way that nations choose to face it. It is of utmost importance that you are familiar with your policies regarding this issue because of this.

At the end of the day, from my point of view, MUN conferences are about learning new things, meeting new people, and having fun. I look forward to our debate and, of course, to meeting you all, albeit through a computer. If you have any questions about the guide or anything else about the topic, do not hesitate to contact me on my email (nick.48478@gmail.com).

Kind regards,

Nick

TOPIC INTRODUCTION

The field of finance generates an asinine amount of data; hundreds of millions of financial transactions globally generate quadrillions of bytes of data that are valuable to all the factors of production in both local and global markets. The sheer amount of data that needs to be stored and processed can prove problematic in a number of ways such as cost, workforce, and a lot more[1].

Big Data in the modern sense is one of the most recent advancements of information categorization in the financial sector because of its ability to analyze, process and store information. Interestingly enough, there is no concrete definition for Big Data; it is, to an extent, subjective. Since there is not a certain threshold that can be considered the lowest floor at which a data set can be considered big data, the definition depends on the perspective of its users. In other words, the definition of Big Data is relative to the size of the business or government that looks to use it[2].

While the concept of Big Data is relatively new, the market for it has managed to create an interesting dynamic when it comes to how it affects the surrounding markets. Such effects include rapid technological advancement in multiple fields, the cultivation of competition as well as the ever-increasing pressure from governments and corporations alike to conform to as well as shape regulations regarding the field of data analytics using Big Data.

The integration of Big Data is a pressing issue for the international community. It is inherently global in nature since the connection between local and international networks is mandatory for favorable results when it comes to processing and analyzing. Big Data also has a sizable intersection with other emerging fields in technology and artificial intelligence, such as the Internet of Things, which can aid in the constant provision of data that corporations can use. Moreover, this issue fits the theme of this year's PSMUN Conference "Redefining Solidarity" quite well, seeing as big data is an excellent way to facilitate satisfactory financial decisions. Furthermore, given the fact that more than 90% of the world's data has been generated after 2014, it is easy to see how the implementation of Big Data, or rather the relative lack, thereof, is a pressing matter.

[1] Hasan, Md. Morshadul, et al. 'Current Landscape and Influence of Big Data on Finance'. *Journal of Big Data*, vol. 7, no. 1, Mar. 2020, p. 21. *BioMed Central*, https://doi.org/10.1186/s40537-020-00291-z

[2] 'Big Data in the Financial Services Industry - From Data to Insights'. *Finextra Research*, 9 Sept. 2019, https://www.finextra.com/blogposting/17847/big-data-in-the-financial-services-industry---from-data-to-insights.

DEFINITION OF KEY TERMS

Big Data

A field of computer science that specializes in analyzing, processing, and storing extremely large data sets in a relatively short period of time.[3]

The IT infrastructure that hosts "Big Data"is called Big Data infrastructure. It includes the tools that collect the data, the software and hardware that stores it, the software or application environments that break it down and analyze it, and the archive infrastructure that stores it after analysis.[4]

Application Programming Interface (API)

It is a software intermediary that allows two or more applications to communicate. In the context of finance, banks usually use private APIs to track and manage their customers' data.[5]

Internet of Things (IoT)

The Internet of Things (IoT) is a network of physical objects that are embedded with sensors or software that are designed to connect and exchange data with other devices and systems over the internet.[6]

Open Banking

Financial technology refers to the use of open APIs that enable third-party developers to create software and services for a financial institution.[7]

Machine learning

It is a field of study in computer algorithms that can gradually improve with data and the passage of time. In this context, it is a method of data analysis that automates analytical model building through artificial intelligence.[8]

Compound Annual Growth Rate (CAGR)

It is the rate of return that would be required for an investment to grow from its beginning balance to its ending balance, assuming the profits were reinvested at the end of each period of the investment's life span.[9]

---

[3] *What Is Big Data? | Oracle*. https://www.oracle.com/big-data/what-is-big-data/

[4] Big Data Infrastructure: 4 Pain Points and How to Solve Them'. *Precisely*, 19 Dec. 2019, https://www.precisely.com/blog/big-data/4-big-data-infrastructure-points-solve

[5] '5 Examples of APIs We Use in Our Everyday Lives | Nordic APIs |'. *Nordic APIs*, 10 Dec. 2019, https://nordicapis.com/5-examples-of-apis-we-use-in-our-everyday-lives/

[6] *What Is the Internet of Things (IoT)?* https://www.oracle.com/internet-of-things/what-is-iot/

[7] 'Open Banking Definition'. *Investopedia*, https://www.investopedia.com/terms/o/open-banking.asp

[8] *Machine Learning: What It Is and Why It Matters*. https://www.sas.com/en_in/insights/analytics/machine-learning.html

[9] 'Compound Annual Growth Rate (CAGR)'. *Investopedia*, https://www.investopedia.com/terms/c/cagr.asp

Cartel

A collection of independent businesses or organizations that collude in order to manipulate the price of a product or a service.[10]

Moore's Law

The principle is that the number of transistors in a microchip doubles every two years though the cost of computers halves. This was an observation made by the co-founder of Intel, Gordon Earle Moore.[11] This law has already reached its limit because there is a physical limit to how many transistors can fit on a silicon chip.

Cybersecurity

The practice of protecting systems and networks from digital attacks that are usually aimed at accessing information with the intent to extort, disrupt or declassify.[12]

BACKGROUND INFORMATION

A Brief historical recursion into statistical analysis and computing

One of the first advancements in the field of analytics was in 1663, when John Graunt, an English statistician and demographer, developed early statistical and census methods that provided a base for modern demography. He is considered by many to be the father of statistical data analysis.[13]

In 1959, Herman Hollerith invented the punch card tabulating the machine. It was designed to store and summarize information stored on punched cards using electromechanics, which is largely related to modern digital machines. During the same time period, Arthur Samuel would be the first to coin the term machine learning. Only a decade later we would see what is largely considered the first network; the Advanced Research Projects Agency Network, otherwise known as ARPANET.[14]

As computers continued to advance from that point onwards another great milestone would be reached in 2006. The creation of Apache Hadoop by Doug Cutting and Mike Cafarella. It is an open-source API that is used to store and process large data sets; the kind that can be seen in Big Data.[15] Not too long after, in 2012, the USA would recognize the innovation and potential. The Obama Administration invested US$200

[10] 'Cartel'. *Investopedia*, https://www.investopedia.com/terms/c/cartel.asp.

[11] 'Moore's Law Explained'. *Investopedia*, https://www.investopedia.com/terms/m/mooreslaw.asp.

[12] 'What Is Cybersecurity?' *Cisco*, https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html.

[13] *John Graunt | English Statistician | Britannica*. https://www.britannica.com/biography/John-Graunt

[14] 'A History and Timeline of Big Data'. *WhatIs.Com*, https://whatis.techtarget.com/feature/A-history-and-timeline-of-big-data

[15] 'An Introduction to Apache Hadoop for Big Data' 26 Aug 2014 Sachin P. Bappalige . *Opensource.Com*, https://opensource.com/life/14/8/intro-apache-hadoop-big-data

million into the Big Data Research and Development Initiative in hopes of developing the wellness and efficacy of all the sectors of the US economy through Big Data.[16] Around the same period, many other nations invested similar amounts.

The market for Big Data would continue to grow at an exponential rate. In 2019, Allied Market Research announced that the Big Data analytics market was worth US$193.14 billion that year alone. Big Data's growth differs from source to source, since there are different samples involved. However, the general consensus is that the market will continue to grow. It is estimated that the market will balloon to approximately US$234 billion dollars by 2026.[17]Other sources suggest much bigger numbers. Allied Market Research estimates that the market will be worth US$420 billion.

Functions of Big Data Infrastructure

Mentioned in the Definition of Key Terms section of this guide, Big Data infrastructure is the physical IT infrastructure that hosts Big Data's functions. Included are the tools that collect data from various sources such as but not limited to the internet and surveys. Also included is the software and hardware that break down and analyze said data, in order for human managers to use the now digestible results and the archive infrastructure that takes all of the aforementioned data and stores it for future reference. All of the above can be customized to an organization's needs, but this main structure remains unchanged.

While conventionally expensive, this structure has a multitude of functions ranging from the storage and analysis of financial records, providing support for financial and business decisions to sharing information with other businesses around the world, making it very versatile. When talking about the financial services sector specifically, Big Data can be used for investment appraisal in the form of almost instantaneous calculation and estimation of revenue streams based on concurrent investments of similar nature and data collected by surveys regarding, for example, the public's response.

Big Data in the financial sector

Big Data's role in the economy is quite pivotal, as data analytics act as the way that corporations make decisions. Evolving the way one analyzes and stores data, no matter the field, is sure to usher in a wave of evolution in the market, directly or indirectly. Besides understanding the nature of Big Data, itself, one must look at its effects on the system it is to be implemented in to gauge whether it is worthwhile.
One of the more obvious changes that the financial sector is sure to undergo is rapid technological evolution in multiple areas. As technology improves, the ways in which

[16] 'PRESS RELEASE: Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments'. *Whitehouse.Gov*, 29 Mar. 2012, https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200

[17] Inc, Global Industry Analysts. *Global Big Data Market to Reach $234.6 Billion by 2026*. https://www.prnewswire.com/news-releases/global-big-data-market-to-reach-234-6-billion-by-2026--301322252.html

data can be collected become more efficient, thus being able to "produce" more data that can prove useful to businesses and governments. There is, therefore, a need for systems that offer high processing power and a large amount of storage, both of which can be found in Big Data Infrastructure.

The rise of open APIs further aids in the increase in demand for Big Data, as more and more parties have access to an increasing amount of information that, again, is useful to them. Open APIs and their intersection with Big Data is a big reason why the latter is so useful in today's financial climate, and as open API technology continues to evolve into more secure and interconnected software, the demand for Big Data is sure not to disappear. To elaborate, open APIs in the field of finance essentially allow institutions to "borrow" data from one another, which fosters competition, and therefore, growth.

The competition between institutions that are getting into or already have Big Data infrastructure is imperative to how this market will evolve. Stakeholders that are able to effectively utilize big data will have the edge over their competitors due to the fact that it is able to provide, for example, advice on financial decisions based on empirical data much faster than its predecessors. New services and products could be created as a result of this competition, and as technology advances, governments and institutions that refrain from investing in Big Data infrastructure will be left "a step behind".

On a related note, newly established regulatory pressure such as the FRTB - the Fundamental Review of the Trading Book - and others, force banks to disclose more data about their operations and clients to governments. Big Data is an effective tool that many of them consider and implement to both conform to the regulations and improve their operations in ways that are mentioned above.

Advantages of big data implementation

Better decision-making

In a survey conducted by the New Vantage partners, 36.2 percent of respondents reported that improved decision making was their flagship goal when it came to their Big Data analytics investments. Furthermore, 84.1 percent had made progress toward that goal, and approximately 59 percent had experienced measurable success, for an overall success rate of 69%. The results of the survey are able to represent the relative success that the use of Big Data analytics has seen over the course of its implementation in multiple sectors of the economy. It can give businesses the data-driven insights they need to help their companies compete and grow.

The growth can be attributed to the fact that big data analytics are able to provide a constant stream of relevant information in an easily digestible form so that the executives responsible are able to make informed decisions. [18]

Reduced costs

Big Data can be used to cut costs through different methods that can be modified to conform to one's needs. One such method is analyzing different management options from different companies to be able to optimize management for maximum operational efficiency. Another way Big Data could help reduce costs is that by virtue of it being an excellent decision-making tool, it could be used to assess how decisions could be made based on prospective costs and profits. [19]

Surge of investment

Big Data is expected to and is already experiencing a monumental surge of investment. Immense companies such as Amazon, Alphabet, and Microsoft have already invested in Big Data infrastructure for obvious reasons. Besides that, the market for Big Data is growing at exponential rates. The projected market value for Big Data is an astonishing ~US$685 billion, which is more than triple the total market value in 2021 (~US$203 billion). This makes sense, as, by 2030, humanity will have produced more than 550 Zettabytes of data (roughly equal to 572 billion terabytes), which will have to be stored, and with Moore's law continuing to have an effect on chip manufacturing, we do not seem to be slowing down when it comes to data generation.[20]

Improvement in Business Functions

Big Data can help businesses and governments alike to grow and increase their survivability. It can help with accounting, development of services or products, whether that would be improving upon existing ones or creating new ones entirely, management and more. It can also help with developing the tertiary sector of production, which focuses on service provision by analyzing what the population demands.[21]

To summarize, Big Data's functions can be utilized in a multitude of ways within an organization to aid in its smooth operation. As should be apparent so far, the ways in which Big Data can be used are endless, especially when it comes to a sector that largely depends on data and statistical analysis like finance. For

---

[18] '7 Benefits to Using Big Data for Small Businesses'. *IndustriusCFO*, 7 Apr. 2015, https://www.industriuscfo.com/7-benefits-using-big-data/

[19] Schleier-Keller, Diane. 'Top 3 Reasons to Invest in Big Data'. *InsideBIGDATA*, 23 Aug. 2017, https://insidebigdata.com/2017/08/23/top-3-reasons-invest-big-data

[20] 'Big Data Pros and Cons'. *Datamation*, 9 Aug. 2018, https://www.datamation.com/big-data/big-data-pros-and-cons/

[21] 'What Is Big Data and What Are Its Benefits?' *Simplilearn.Com*, https://www.simplilearn.com/tutorials/big-data-tutorial/what-is-big-data

this reason, it is an invaluable tool for development, whose merit is noteworthy to say the least.

Disadvantages of Big Data

Need for specialized personnel

Big Data infrastructure requires a lot of human resources to set up and run. To start, depending on the scale of the project, construction crews may be needed to create a space for the infrastructure. The logistics for bringing components from one end to the world is very complicated and finally, maintenance and cleaning of the spaces is very important too, especially when it comes to the more sensitive components.

One may have noticed that all the personnel mentioned above do not need any special training or expertise in a specific field. However, Big Data analytics and infrastructure require a variety of specialized personnel. Analytics, IT, hardware engineering specialized maintenance are just some of the many oddly specific professionals that are needed when looking to operate a system of this scale.

Varying Data Quality

Not all data is created equal. Data quality is relative to the data's intended use and the individual or individuals looking to analyze it. To elaborate, consumers are very sensitive when it comes to changes in their environment and moods, therefore, data that is collected in the summer, for example, may be different than data that is collected in the winter. Moreover, due to the vast nature of the data, it is difficult to pick and choose the parameters for the data one needs, which may lead to inconsistent results upon analysis.

Compliance with intergovernmental regulations

Complying with intergovernmental regulations regarding personal privacy is very difficult, especially when it comes to the financial sector. International laws such as the European Union's GDPR limit the kind of data that businesses are able to share with each other since a lot of the information is personally identifiable and, therefore, quite sensitive. Sharing such information could lead to breaches of privacy.

Furthermore, sharing too much information could lead to an indirect breach of privacy. Combining multiple data sets may, in some cases, lead to the resulting processed information being identifiable, meaning that it can be traced back to individuals or groups, subjecting them to potential hazards.

Cybersecurity

One of the most prominent and threatening problems that come with Big Data is the ever-present threat of cyberattacks. A well-coordinated and executed cyberattack on a Big Data system that houses and processes all the data in an

area could be catastrophic, not only for the local government but also for all the people living in that area. This is particularly burdensome for the reason that data centers that store and process financial data are very lucrative targets for cyber attackers since the potential payout is huge for just one "job".

Extended hardware needs

Big Data infrastructure requires a lot of hardware to store the immense amounts of data that the infrastructure is to be built for. Big Data hardware needs can be split into two categories: storage and processing.

Storage requires an exponential amount of space and a lot of maintenance. Storage usually comes in a Redundant Array of Independent Disks (RAID) which means that a lot of disks work as one, allocating data between them evenly, depending on the methods RAID is implemented with. This system is usually quite reliable, and, as the name suggests, redundant. However, it is costly to maintain and, nevertheless, requires a lot of actual hardware to sustain the data that needs to actually be stored.

Processing does not require a lot of space, but it is very expensive and tricky to work with. Not all processors (CPUs) are suitable for this kind of work, and depending on the needs of the stakeholders, there may be a need for more than CPUs. Cooling and organization are big problems too.

Large amounts of processing and storage are not only required for any Big Data infrastructure project but they are also very expensive. Due to the global microchip shortage, both processors and solid-state drives (Drives using flash memory instead of mechanical discs) are experiencing an increase in prices, which is extremely unfavorable for organizations or people looking to obtain large quantities of both.

High Initial Cost

Big Data expansion projects almost always go over budget according to Datamation.[22] Although open-source software brings down costs dramatically, the bulk of the spendings stems from hardware and staffing needs, a lot of which is needed as mentioned above. Of the two, hardware is the costliest.

The cost of Big Data platforms such as Hadoop and Spark scales up proportionally to the amount of storage and processing power that the business uses. A single Hadoop cluster can be composed of anywhere from one node to a virtually infinite number of nodes. Each of these clusters is recommended to have at least a mid-range server for the processing which costs about $1000-$2000 per terabyte. Therefore, a petabyte Hadoop cluster will cost approximately $1 million since it needs about 200 nodes. This cost may seem relatively insignificant, but when one considers the asinine amounts

[22] *'Big Data Pros and Cons'. Datamation, 9 Aug. 2018, https://www.datamation.com/big-data/big-data-pros-and-cons/*

of data that usually comprise a Big Data set, a lot more than a petabyte of storage is needed to facilitate effective storage and processing.

Correlation errors

A common technique used in data analysis is to draw correlations between two data sets. The aforementioned data sets have to be correlated in the real world for the analysis to be effective. If they, in fact, are not related in the real world, then correlation errors occur, which can lead to wrong financial decisions. For example, when analyzing two data sets, such as ice cream sales and homicides, one may see that their patterns of increases and decreases somewhat correlate, which may lead to the false assumption that one is causation for the other. This can be misleading for a decision-making part of a business.

To summarize, Big Data is a really diverse field that offers a lot of advantages, especially in the field of finance. It is critical that measures of international magnitude are implemented to facilitate sustainable development through Big Data while also ensuring the safety of its users and the local and in some cases global[23] economy.

MAJOR COUNTRIES AND ORGANIZATIONS INVOLVED

United States of America (USA)

With a staggering 51% market share in Big Data analytics according to Statista, the USA is by far the largest stakeholder in the market, with the immediate follower being Japan, at 5.7% market share. The staggering difference has to do with the increase in data demands in sectors such as public utilities.[24] In 2012, the Obama Administration invested US$200 million in research and development initiatives for Big Data – The "Big Data Research and Development Initiative". Its goals were to "advance state-of-the-art core technologies needed to collect, store preserve, manage, analyze, and share huge quantities of data", to "harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning" and to "expand the workforce needed to develop and use Big Data technologies".

Big government agencies such as the National Science Foundation and Department of Defense are part of the initiative, with the latter committing to investing

[23] In global banking and logistics markets.

[24] *Top 10 Countries & Regions Leading the Big Data Adoption in 2019*. 24 Nov. 2019, https://www.analyticsinsight.net/top-10-countries-regions-leading-the-big-data-adoption-in-2019/, 'Big Data & Analytics Market Share by Country 2021'. *Statista*, https://www.statista.com/statistics/1258046/worldwide-big-data-business-analytics-market-share-by-country

approximately $250 million annually across the Military Departments in hopes of using the initiative to improve the effectiveness of the military as a whole through better decision making and provision of support. In addition, the National Institute of Health announced that the then-largest data set on human genetics – the 1000 Genomes Project – would become publicly available on Amazon Web Services. The 200-terabyte project is a prime example of the use of Big Data, especially in 2012. Compared to today's standards for Big Data, the project's size seems insignificant.[25]

People's Republic of China

In a report from global market research firm International Data Corp, China's Big Data market share was estimated to exceed US$10 billion for the first time in 2020. Around 38% of that was spent on banking, telecommunication, and local governments. It has been established that China's Big Data market will witness an upward trend as fintech innovation and digital government construction fuel the continuous increase of industry users' demand for Big Data. According to the analyses, China's Big Data market is predicted to hit $20 billion by 2024, marking a 145 percent growth from the rate recorded in 2019.[26]

United Kingdom (UK)

The UK's Big Data Analytics market value, according to Statista, exceeded US$3.6 billion dollars and is expected to grow at an exponential rate from 2021 onwards.[27] The UK also invested 158 million pounds sterling in 2011 and 375 million in 2012, making it one of the first UN member states to invest in Big Data infrastructure. As of now, there are approximately 670 businesses in the UK that pertain to Big Data and Big Data infrastructure according to CrunchBase.

Japan

Japan, being the second largest Big Data Analytics shareholder, is likely to see steady expansion akin to other developed nations already invested in Big Data, such as the USA and China. [28]In a survey conducted by the European patent office in 2020, Japan was the second-largest country by patent applications relating to advanced IT fields such as artificial intelligence. [29]

---

[25] 'PRESS RELEASE: Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments'. *Whitehouse.Gov*, 29 Mar. 2012, https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200

[26] 孙迟. *China's Big Data Market to Continue Expansion: Report*. https://www.chinadaily.com.cn/a/202103/15/WS604ef289a31024ad0baaf3b3.html

[27] 'Big Data & Analytics Market Share by Country 2021'. *Statista*, https://www.statista.com/statistics/1258046/worldwide-big-data-business-analytics-market-share-by-country, *Top 10 Countries & Regions Leading the Big Data Adoption in 2019*. 24 Nov. 2019, https://www.analyticsinsight.net/top-10-countries-regions-leading-the-big-data-adoption-in-2019/

[28] *Top 10 Countries & Regions Leading the Big Data Adoption in 2019*. 24 Nov. 2019, https://www.analyticsinsight.net/top-10-countries-regions-leading-the-big-data-adoption-in-2019/0

[29] 'Japan Comes in 2nd Worldwide in Number of AI and Big Data Applications'. *Lexology*, 1 Feb. 2021, https://www.lexology.com/library/detail.aspx?g=08273de9-b847-4a3e-8dcd-9831e120e1a9

Russian Federation

The main consumers of Big Data technologies here are banks, telecommunication, and large retailers, with the market value expected to reach 300 billion robles by 2024. The government also plans for implementations of Big Data in the public sector, which can be used by federal agencies as a resource for development. [30]More recently, in July of 2021, Russia approved its first national standard in the field of Big Data. The GOST (an acronym for государственный стандарт, meaning state standard) mainly aims to sync already existing Big Data initiatives within Russia to international ones to facilitate cooperation and communication and to avoid confusion.

TIMELINE OF EVENTS

DATE	DESCRIPTION OF EVENT
1663	John Graunt invents statistical data analysis to combat the bubonic plague.
1884	Herman Hollerith invents the punch card tabulating the machine which was the beginning of data processing.
1943	The UK creates the Colossus, which was designed to decipher Nazi codes in WW2. It performed Boolean and counting operations to analyze large amounts of data.
1959	Arthur Samuel coins the term Machine learning.
1965	The USA plan the first-ever data center to store tax returns and fingerprints on magnetic tape.
October 29th, 1969	Advanced Research Projects Agency Network (ARPANET) is created.
1996	Digital storage becomes cheaper than paper according to IBM Systems Journal paper which was published in 2003.
March 29th, 2012	The Obama administration invests $200 million into the Big Data Research and Development Initiative.
2013	The global market for big data is worth US$10 billion.
2021	Allied Market Research announces that the Big Data Analytics market was worth US$193.14 billion in 2019 and is expected to grow to US$420 billion at a CAGR of 10.9%.

[30] *Top 10 Countries & Regions Leading the Big Data Adoption in 2019*. 24 Nov. 2019, https://www.analyticsinsight.net/top-10-countries-regions-leading-the-big-data-adoption-in-2019/0

PREVIOUS ATTEMPTS TO SOLVE THE ISSUE

Private Businesses and Big Data

A number of large corporations are already investing in their own unique Big Data infrastructure solutions. The Starbucks Corporation uses Big Data in a variety of ways, the most prevalent being personalized promotions, which has been done since 2017 with the "Digital Flywheel" program, which incorporates four pillars (rewards, personalization, payment, order).[31] Another type of company that uses Big Data is streaming platforms such as Spotify and Netflix. Both being very successful services, they not only use Big Data for storing and streaming content on their platforms but also for personalized recommendations based on any given customer's viewing or listening habits, creating a more familiar and streamlined experience. All of these examples demonstrate the efficacy of Big Data infrastructure in businesses that are not only centered around the provision of services but also more traditional, so to speak, businesses.

The Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG)

On August 29th, 2014 the then Secretary-General of the United Nations, Ban Ki-moon named an Independent Expert Advisory Group on a Data Revolution for Sustainable Development to provide information that could be used to shape "an ambitious and achievable vision" for the future of data. The IEAG consists of more than 20 internationally recognized experts in the field of data. The subsequent report highlights two big global concerns with the state of data back in 2014; the challenge of gaps in what we know from data and the challenge of the inequality between those who have data and those who do not.

The report also addresses in detail various issues mentioned in this guide. Examples include the issue of varying data quality and the need for immediate action in both the private and the public sectors.[32]

UN Global Pulse

The UN Global Pulse is an innovation initiative by the UN that was founded in 2009. It aims to "bring real-time monitoring and prediction to development and aid programs" in the form of high-impact analysis solutions for UN member and observer states. It aims to achieve this through its network of data science innovation centers dubbed Pulse Labs that are located in Jakarta, Kampala, and the UN Headquarters in New York City. The initiative was to an extent successful, having created measures that safeguard the privacy and drive innovation in the field of Big Data.[33]

---

[31] Pathak, Ritesh. *6 Ways in Which Starbucks Uses Big Data | Analytics Steps*. https://www.analyticssteps.com/blogs/6-ways-which-starbucks-uses-big-data

[32] 'UN Data Revolution'. *UN Data Revolution*, https://www.undatarevolution.org/

[33] *UN Global Pulse - Big Data for Development and Humanitarian Action*. 31 May 2018, https://www.unglobalpulse.org/

POSSIBLE SOLUTIONS

It is impossible to think about finding some possible solutions to the issue without addressing the drawbacks of Big Data. The solutions mentioned below are targeted towards addressing the problems with Big Data and are recommended. However, there is room for finding ways in which governments can use Big Data analytics to better their circumstances and collaborate with other nations to further develop the economy and financial sector.

Establishing Data Fidelity

As mentioned beforehand, a lack of data fidelity is a real issue that comes with Big Data due to the fact that the scope of some of the collection methods is just too wide. One of the things that can be done in narrowing down the scope of the research or collection of data by limiting which sensor groups will collect data at any given time for example. While it may work in some cases, it would mean that data is susceptible to bias, or simply is too little to make investments in Big Data infrastructure justifiable in others. For that reason, ways to increase data fidelity and validity without narrowing its scope are imperative to the evolution of Big Data.

Data filters

There are multiple solutions to this problem, but they all come with their own problems. One of the solutions that are most commonly found are data filters that take data that is false or not needed out of the settings automatically. It is a relatively cheap way to mitigate some of the problems that can be implemented in a variety of ways, such as simple if/if not-type APIs or even machine learning. On the other hand, they are not perfect when it comes to filtering information that involves handwriting or text written by humans. In addition, they create bottlenecks when implemented in very large data sets, because of the fact that all the data that is to be processed has to go through the filter.

Manual inspection

Another solution to the same problem would be to have employees manually check the data used for any given decision. This would almost completely eliminate error but would mean that the cost to check every single piece of data within a data set big enough to be considered Big Data comes with an elevated cost, to say the least. This solution could be used in cases where the data collection scope is relatively low and the amount of data that needs to be filtered is relatively small.

Machine learning

A middle ground of sorts to the two solutions would be machine learning. It is an effective way to sort through data. As mentioned in the Key Terms section of this guide, machine learning is a field in computer science that specializes in developing artificial intelligence systems that are able to improve with the input of data and time. This is advantageous for Big Data infrastructure as the data input is sufficient, to say the least, for the system to develop while being relatively inexpensive to implement.

It, however, may need support both monetarily and from personnel in the early stages of development or implementation.

Addressing cybersecurity concerns

Big Data is a tool that is able to aid in the mitigation of cybersecurity threats, even though it is a potential target itself. As mentioned before, financial institutions that have Big Data infrastructure in place are very potentially lucrative for cyberattacks because of the sheer density of information, especially because of the fact that it pertains to money. An effective, well-timed, and quick attack would be detrimental to the economy of a nation, as can be seen in the 2016 Bangladesh Bank Heist. Therefore, cybersecurity is at the top of the list when it comes to solutions. In the world of cybersecurity, there are a lot of ways to prevent data loss.

Filters to test and isolate phishing links, trojans and other malicious software that fall into the same category are essential here. The example of the Bangladesh Bank Heist can be used here as well, as the hackers gained access to the bank's network through an infected file that was opened on an employee's personal email address. Therefore, avoiding trojans and phishing links requires some awareness training for the employees responsible. Another way to combat cybersecurity is through the use of decentralized data centers, which essentially break up the infrastructure into smaller bits making the whole harder to attack. This would mean that the cost would be drastically increased of course, but the increased security would be worth it to some.

In general, cybersecurity is something every nation should be concerned about, especially when it comes to the big investments that come with Big Data infrastructure.

BIBLIOGRAPHY

Hasan, Md. Morshadul, et al. 'Current Landscape and Influence of Big Data on Finance'. *Journal of Big Data*, vol. 7, no. 1, Mar. 2020, p. 21. *BioMed Central*, https://doi.org/10.1186/s40537-020-00291-z

'Big Data in the Financial Services Industry - From Data to Insights'. *Finextra Research*, 9 Sept. 2019, https://www.finextra.com/blogposting/17847/big-data-in-the-financial-services-industry---from-data-to-insights.

What Is Big Data? | Oracle.https://www.oracle.com/big-data/what-is-big-data/

Big Data Infrastructure: 4 Pain Points and How to Solve Them'. *Precisely*, 19 Dec. 2019, https://www.precisely.com/blog/big-data/4-big-data-infrastructure-points-solve

'5 Examples of APIs We Use in Our Everyday Lives | Nordic APIs |'. *Nordic APIs*, 10 Dec. 2019, https://nordicapis.com/5-examples-of-apis-we-use-in-our-everyday-lives/

What Is the Internet of Things (IoT)? https://www.oracle.com/internet-of-things/what-is-iot/

'Open Banking Definition'. *Investopedia*, https://www.investopedia.com/terms/o/open-banking.asp

Machine Learning: What It Is and Why It Matters. https://www.sas.com/en_in/insights/analytics/machine-learning.html

'Compound Annual Growth Rate (CAGR)'. *Investopedia*, https://www.investopedia.com/terms/c/cagr.asp

'Moore's Law Explained'. *Investopedia*, https://www.investopedia.com/terms/m/mooreslaw.asp.

'7 Benefits to Using Big Data for Small Businesses'. *IndustriusCFO*, 7 Apr. 2015, https://www.industriuscfo.com/7-benefits-using-big-data/

Schleier-Keller, Diane. 'Top 3 Reasons to Invest in Big Data'. *InsideBIGDATA*, 23 Aug. 2017, https://insidebigdata.com/2017/08/23/top-3-reasons-invest-big-data/

'Big Data Pros and Cons'. *Datamation*, 9 Aug. 2018, https://www.datamation.com/big-data/big-data-pros-and-cons/

Top 10 Countries & Regions Leading the Big Data Adoption in 2019. 24 Nov. 2019, https://www.analyticsinsight.net/top-10-countries-regions-leading-the-big-data-adoption-in-2019/

'Big Data & Analytics Market Share by Country 2021'. *Statista*, https://www.statista.com/statistics/1258046/worldwide-big-data-business-analytics-market-share-by-country

孙迟. *China's Big Data Market to Continue Expansion: Report*. https://www.chinadaily.com.cn/a/202103/15/WS604ef289a31024ad0baaf3b3.html

Pathak, Ritesh. *6 Ways in Which Starbucks Uses Big Data | Analytics Steps*. https://www.analyticssteps.com/blogs/6-ways-which-starbucks-uses-big-data

'PRESS RELEASE: Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments'. *Whitehouse.Gov*, 29 Mar. 2012, https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200

Hajirahimova, Makrufa Sh., Alilyeva, Aybeniza S. "BIG DATA INITIATIVES OF DEVELOPED COUNTRIES", "Outlier detection on Big Data", January 2017, https://www.researchgate.net/publication/317486245_BIG_DATA_INITIATIVES_OF_DEVELOPED_COUNTRIES

'What Is Cybersecurity?' *Cisco*, https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html.

'Japan Comes in 2nd Worldwide in Number of AI and Big Data Applications'. *Lexology*, 1 Feb. 2021, https://www.lexology.com/library/detail.aspx?g=08273de9-b847-4a3e-8dcd-9831e120e1a9.

UN Global Pulse - Big Data for Development and Humanitarian Action. 31 May 2018, https://www.unglobalpulse.org/
'UN Data Revolution'. *UN Data Revolution*, https://www.undatarevolution.org/

'An Introduction to Apache Hadoop for Big Data' 26 Aug 2014 Sachin P. Bappalige . *Opensource.Com*, https://opensource.com/life/14/8/intro-apache-hadoop-big-data

John Graunt | English Statistician | Britannica. https://www.britannica.com/biography/John-Graunt

Inc, Global Industry Analysts. *Global Big Data Market to Reach $234.6 Billion by 2026*. https://www.prnewswire.com/news-releases/global-big-data-market-to-reach-234-6-billion-by-2026--301322252.html